

Punishment-induced behavioral and neurophysiological variability reveals dopamine-dependent selection of kinematic movement parameters

Galea, Joseph M; Ruge, Diane; Buijink, Arthur; Bestmann, Sven; Rothwell, John C

DOI:

[10.1523/JNEUROSCI.1294-12.2013](https://doi.org/10.1523/JNEUROSCI.1294-12.2013)

License:

None: All rights reserved

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Galea, JM, Ruge, D, Buijink, A, Bestmann, S & Rothwell, JC 2013, 'Punishment-induced behavioral and neurophysiological variability reveals dopamine-dependent selection of kinematic movement parameters', *The Journal of Neuroscience*, vol. 33, no. 9, pp. 3981-8. <https://doi.org/10.1523/JNEUROSCI.1294-12.2013>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Punishment-Induced Behavioral and Neurophysiological Variability Reveals Dopamine-Dependent Selection of Kinematic Movement Parameters

Joseph M. Galea,^{1,2} Diane Ruge,² Arthur Buijink,² Sven Bestmann,² and John C. Rothwell²

¹Behavioural Brain Sciences, School of Psychology, University of Birmingham, B15 2TT United Kingdom, and ²Sobell Department of Motor Neuroscience and Movement Disorders, UCL Institute of Neurology, University College London, WC1N 3BG United Kingdom

Action selection describes the high-level process that selects between competing movements. In animals, behavioral variability is critical for the motor exploration required to select the action that optimizes reward and minimizes cost/punishment and is guided by dopamine (DA). The aim of this study was to test in humans whether low-level movement parameters are affected by punishment and reward in ways similar to high-level action selection. Moreover, we addressed the proposed dependence of behavioral and neurophysiological variability on DA and whether this may underpin the exploration of kinematic parameters. Participants performed an out-and-back index finger movement and were instructed that monetary reward and punishment were based on its maximal acceleration (MA). In fact, the feedback was not contingent on the participant's behavior but predetermined. Blocks highly biased toward punishment were associated with increased MA variability relative to blocks either with reward or without feedback. This increase in behavioral variability was positively correlated with neurophysiological variability, as measured by changes in corticospinal excitability with transcranial magnetic stimulation over the primary motor cortex. Following the administration of a DA antagonist, the variability associated with punishment diminished and the correlation between behavioral and neurophysiological variability no longer existed. Similar changes in variability were not observed when participants executed a predetermined MA, nor did DA influence resting neurophysiological variability. Thus, under conditions of punishment, DA-dependent processes influence the selection of low-level movement parameters. We propose that the enhanced behavioral variability reflects the exploration of kinematic parameters for less punishing, or conversely more rewarding, outcomes.

Introduction

Action selection is often described as the cognitive decision process that selects between competing movements. The concepts of cost and reward (Schultz, 2006) have been used to show that dopamine (DA)-dependent processes select an action that optimizes reward and minimizes cost/punishment (Frank et al., 2004; Pessiglione et al., 2006). When attempting to make such a selection, learning from trial and error is vital (Fee and Goldberg, 2011). For this process to be successful, behavior has to be initially variable so that sufficient task space is explored to attain the desired balance between cost and reward. For example, in mice it has been shown that DA is important for inducing novel activity patterns in cortico-Basal Ganglia circuits that drive such motor exploration (Costa et al., 2006; Costa, 2011). In pigeons, D₂ ago-

nists increase behavioral variability during reinforcement learning (Pesek-Cotton et al., 2011), and work in songbirds suggests that DA shapes neural and behavioral variability by providing a reinforcement signal that indicates good or bad song performance (Fee and Goldberg, 2011). Despite these strong links between DA and behavioral variability in animals, the relationship between behavioral variability, reward/punishment, and DA in humans is relatively unknown.

Recent work shows that once an action has been chosen, DA-dependent processes of selection also influence low-level movement parameters such as movement time and force (Mazzoni et al., 2007; Pessiglione et al., 2007). Specifically, Mazzoni et al., (2007) showed that Parkinson's disease patients who suffer from DA depletion implicitly select to move slower even though they preserve the ability to execute faster movement speeds with similar accuracy. This suggests that in patients there is an abnormal balance between the costs of moving fast and the rewards of completing the task, and that the selection of kinematic parameters is under dopaminergic influence (Mazzoni et al., 2007).

Here we sought to test whether the selection of low-level movement parameters is affected by punishment and reward in ways similar to those that operate during high-level action selection. Specifically, we asked whether variability in kinematic parameters can be influenced by punishment and reward, and whether this is DA dependent. To assess this, participants per-

Received March 16, 2012; revised Dec. 18, 2012; accepted Jan. 15, 2013.

Author contributions: J.M.G. and J.C.R. designed research; J.M.G., D.R., and A.B. performed research; J.M.G. analyzed data; J.M.G., D.R., A.B., S.B., and J.C.R. wrote the paper.

This work was supported by the European Union framework 7 initiative REPLACES (J.M.G., J.C.R.), Birmingham Fellows scheme (J.M.G.), Tourette Syndrome Association (D.R., J.C.R.), the Prinses Beatrix Fonds (A.B.), Biotechnology and Biological Sciences Research Council (S.B.), and European Research Council (ActSelectContext, 260424; S.B.).

Correspondence should be addressed to Joseph M. Galea, School of Psychology, University of Birmingham, Birmingham, B15 2TT United Kingdom. E-mail: joe.galea.01@gmail.com.

DOI:10.1523/JNEUROSCI.1294-12.2013

Copyright © 2013 the authors 0270-6474/13/333981-08\$15.00/0

formed an out-and-back finger movement and were instructed that monetary reward and punishment were based on its maximal acceleration (MA). In fact, the feedback was not contingent on the participant's behavior but predetermined by set probabilities. Thus, in a block highly biased toward punishment, participants were often unable to attain reward but were still required to select a MA. We predicted that this should enhance MA variability as a result of the participant searching for a less punishing or more rewarding outcome. We then tested the effect of manipulating the DA system by giving a selective D₂ antagonist. Furthermore, we asked whether this manipulation also had an impact on the final stage of action selection by measuring the variability of corticospinal excitability (CSE) with transcranial magnetic stimulation (TMS) over the primary motor cortex (M1).

Materials and Methods

Participants. Twenty-four self-assessed, right-handed individuals with no history of neurological or psychiatric conditions (13 women; mean age, 26 ± 7 years old; age range, 19–44 years) participated in the study. The study was approved by the Joint Research Ethics Committee of the National Hospital for Neurology and Neurosurgery and the Institute of Neurology at University College London and was in accordance with declaration of Helsinki. Written, informed consent was obtained from all participants.

General procedure. All experiments were conducted in a double-blind, placebo-controlled, crossover design. Each subject participated in two experimental sessions separated by at least 1 week. For each session, participants received either 400 mg of the D₂ antagonist sulpiride or an equivalent placebo 1.5 h before the onset of the task so that the latter coincided with the peak plasma concentration of sulpiride (Deleu et al., 2002). This is the same dose as in previous studies that has shown clear effects on M1 plasticity protocols (Nitsche et al., 2006; Monte-Silva et al., 2011). A D₂ antagonist was chosen due to previous results showing a specific effect of D₂ receptors on behavioral variability (Pesek-Cotton et al., 2011) and also the D₂ receptor's apparent responsiveness to punishment/negative outcomes (Cools et al., 2009; Kravitz et al., 2012).

At the end of each session, participants reported their attention and fatigue using a self-scored visual analog scale in which 1 represented poorest attention and maximal fatigue and 7 represented maximal attention and least fatigue (Galea et al., 2009).

Experiments 1–3. Twelve subjects (6 women; mean age, 28 ± 8 years old; age range, 20–42 years) participated in experiments 1–3. These experiments were performed in the same two sessions (sulpiride, placebo) with the order of experiments 1 and 2 counter-balanced across participants.

Experiment 1: DA-dependent selection of kinematic parameters. Experiment 1 examined whether DA-dependent processes of selection can influence kinematic parameters such as MA. To this end, we investigated how behavioral and neurophysiological variability were influenced by punishment and reward, and whether these changes were dependent on DA.

The experimental procedure was identical for both sessions (sulpiride, placebo). To assess movement, a one-degree of freedom accelerometer (Entran) was placed on the proximal phalanx of the index finger. Participants were seated in a chair with a computer screen positioned at eye level ~ 30 cm in front of them. They were instructed to place their right-hand on a table so that it was at rest. A single trial was then explained (Fig. 1a). Initially, the participants would see "WAIT" for 2000 ± 200 ms, after which an arrow would appear for 250 ms that pointed to the left. This was the go cue to make an adduction movement with their right index finger (Fig. 1a). Participants were instructed to make an out-and-back horizontal movement that returned the index finger to a resting position. Following 700 ms, a white square appeared either at the bottom (punishment), top (reward), or middle (no change) of the screen for 100 ms. A single TMS pulse (see below, Assessment of CSE excitability) was then given at 150 ms after the onset of the square appearing. The timing of the TMS pulses were thought to capture the response of phasic DA to

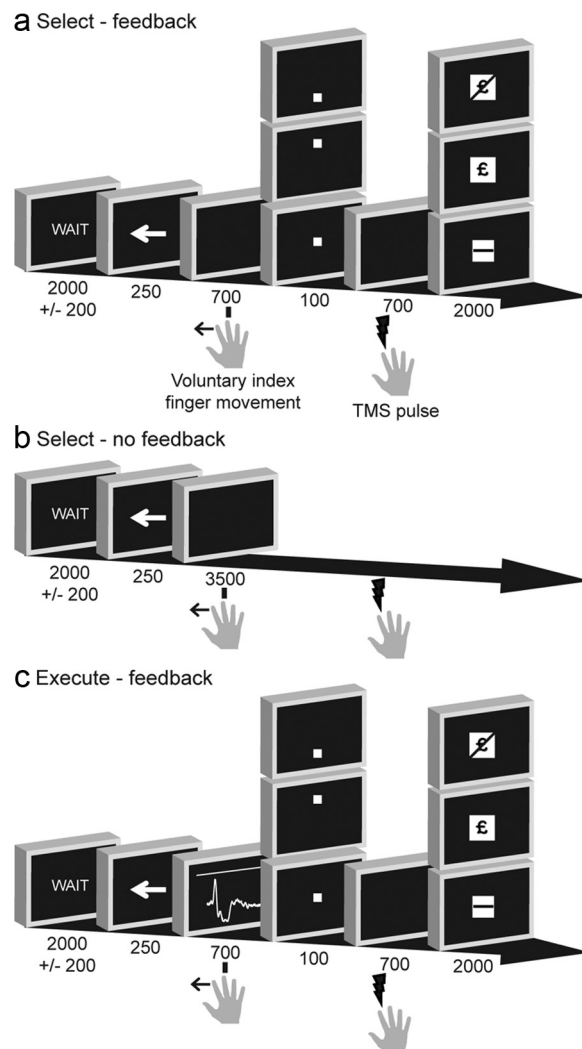


Figure 1. Experimental design. *a*, Select—feedback. A “WAIT” signal was followed by a go cue (leftward arrow). This was the signal for participants to make an adduction movement with their index finger. Following 700 ms, a white square appeared either at the bottom (punishment), top (reward), or middle (no change) of the screen. A single TMS pulse was then given at either 150 or 300 ms after the onset of the square appearing. After another 700 ms, a £ sign with a crossbar through it (punishment), a £ sign (reward), or a horizontal line (no change) appeared. These symbols related to monetary punishment, reward, or no change. Participants were instructed that the feedback was a result of their movement MA when in fact it was predetermined by set probabilities. *b*, Select—no feedback. Participants made a similar adduction movement with their index finger, however they received no feedback. A single TMS pulse was administered at a time point that was equivalent to the 150 ms pulse in experiment 1. *c*, Execute—feedback. Participants were able to view their acceleration online and were instructed to execute a specific MA as indicated by a horizontal line on the screen. They were told that the monetary reward and punishment was based on their ability to execute the specific MA when in fact it was once again predetermined. A single TMS pulse was administered at a time point that was equivalent to the 150 ms pulse in *a*. Participants also had to perform this execution protocol without punishment/reward feedback (no feedback; see *b*).

visual stimuli (Redgrave and Gurney, 2006; Schultz, 2007). After another 700 ms, a £ symbol with a crossbar through it, a £ symbol, or a horizontal line appeared for 2000 ms (Fig. 1a). These symbols related to monetary punishment, reward, or no change, respectively. Importantly, there was a fixed relationship between the position of the white box and type of feedback the participants received (Fig. 1a). This meant that participants could learn these associations and use the position of the box to predict the feedback. It was decided to create these associations because phasic DA is driven by the position of visual stimuli, whereas less is known regarding DA's roles during the process of object recognition required with feedback (Redgrave and Gurney, 2006).

There were 3 blocks of 50 trials with monetary feedback. Participants started the experiment with £10. On each trial with reward feedback participants received £0.25, while punishment feedback would cause them to lose £0.25. No change feedback meant the level of money remained unchanged. At the end of each block, the amount of money participants received would either be added to or subtracted from their initial £10. Participants were instructed that the feedback they received was a consequence of their outward movement's MA, with the optimal MA possibly being either fast or slow. Unbeknown to participants, the feedback was probabilistic and not a consequence of their own movement. The first block was used to familiarize the participants to the experiment. All 3 forms of feedback had an equal probability (0.33) of occurring. This allowed participants to become aware that the position of the white boxes predicted the forthcoming feedback. The following 2 blocks consisted of either reward or punishment occurring with a probability of 0.8, with the other 2 forms of feedback having a probability of 0.1. The probabilistic feedback was critical, as it meant participants remained naive as to the feedback's predetermined nature. In addition, the amount of money participants earned across the 3 blocks equaled to £0; however, participants were unaware of the money they had earned until the end of the study.

Baseline performance was assessed by a no feedback block. Within this block (50 trials) participants were instructed to make an out-and-back horizontal index finger movement that was followed by a black screen for 3500 ms (no feedback). Participants were told that the no feedback block was independent of any monetary reward or punishment and to simply select a MA based on their own volition. Nine hundred fifty milliseconds after the onset of the blank screen, a TMS pulse was administered that was equivalent in terms of timing to the 150 ms TMS pulse within the reward/punishment blocks (Fig. 1*b*). No feedback was used rather than the no change feedback described above (Fig. 1*a*), as a block without any visual feedback or relationship to money was thought to be a better indicator of baseline performance. The order of the no feedback, reward, and punishment blocks were counterbalanced across participants.

Following our analysis, we recruited an additional 12 subjects (7 women; mean age, 26 ± 6 years old; age range, 19–44 years) for experiment 1 as an independent confirmation of the main results and additionally to test for the specificity of the time of TMS. These participants were exposed to the training, reward, and punishment blocks (Fig. 1*a*) across placebo and sulpiride sessions. Importantly, the TMS pulse could now occur either 150 or 300 ms after the presentation of the white square. The order of the TMS times was random; however, each was given in 50% of the trials across a block.

Experiment 2: execution of predetermined MA. Movement parameters such as MA are often thought of in terms of action execution (accuracy/precision of a particular MA) rather than action selection. As our aim was to show that DA-dependent processes of selection can influence kinematic variables such as MA, it was important to dissociate selection from execution. Therefore, experiment 2 replicated the training, reward, punishment, and no feedback blocks of experiment 1 with TMS being applied at a similar time point (Fig. 1*c*). However, following the presentation of the “go” cue, participants were now able to observe their acceleration online and told to execute a predetermined MA by attempting to hit a “target line” with their MA (Fig. 1*c*). The MA target was chosen by averaging all MA mean values from experiment 1 ($3.3 \log \text{ m/s}^2$). The participants were informed that reward and punishment were based upon their ability to execute the target MA.

Experiment 3: influence of sulpiride on resting CSE excitability. Three sets of 20 single-pulse TMS measurements (5 s interval between pulses) were recorded to assess CSE excitability at rest (see below, Assessment of CSE excitability). These were performed before (T1), between (T2), and after (T3) experiments 1 and 2.

Assessment of CSE excitability. TMS-elicited motor-evoked potentials (MEPs) were recorded to measure excitability changes of the M1 representation of the task-involved first dorsal interosseous (FDI) and the task-noninvolved adductor digiti minimi brevis (ADM) muscles of the right hand. Single-pulse TMS was applied with a Magstim 200 magnetic stimulator using a figure-eight magnetic coil (diameter of one winding, 70 mm; peak magnetic field, 2.2 T). The coil was held tangentially to the

skull, with the handle pointing backward and laterally at an angle of 45° from midline. The optimal coil position was determined by the location on the scalp where stimulation consistency resulted in the largest MEP at rest for the FDI (“motor hot spot”). During the initial behavioral training block of each session, the TMS intensity was adjusted so that an approximate MEP of 1 mV was attained. This TMS intensity was then used throughout the session. Electromyographic (EMG) recording was made from both the FDI and ADM with Ag-AgCl electrodes in a belly-tendon montage. Responses were amplified with a D360 amplifier (Digitimer) and filtered at 20 Hz and 2 kHz with a sampling rate of 2 kHz. All behavioral and neurophysiological data were recorded using Signal software (Cambridge Electronic Design) and analyzed offline with Matlab (MathWorks).

Data analysis. Behavioral data were associated with the feedback on the preceding trial. For each outward index finger movement, maximal MA (m/s^2) and reaction time (RT; ms) were calculated. RT was measured as the time between the “go” cue (arrow) and acceleration reaching 10% of maximum. Any movement with a RT > 800 ms was removed (<3%). As the task instructions clearly stated that feedback was dependent on MA, we did not expect to observe any manipulation of RT across blocks or sessions.

The MEP response was associated with the feedback received on the current trial. The peak-to-peak MEP amplitude was calculated for the FDI (involved) and ADM (noninvolved) muscles. Preactivation was defined as the averaged rectified EMG activity for the 100 ms before the TMS pulse. Any value above 100 microvolt (Bestmann et al., 2008) resulted in the MEP being removed from analysis (<6%). Finally, MEPs with amplitudes of <0.05 mV were removed as these might represent trials in which an MEP was not actually obtained (<1%). Overall <10% of trials were excluded.

MA and neurophysiological data were multiplied by 1000 and then log transformed. The multiplication ensured the log-transformed data were positive throughout, as any value that is <1 and then log transformed produces a negative number. We believed this would improve the clarity of the results. RT was simply log transformed. For every participant and session, we calculated the mean and within-subject standard deviation for the reward and punishment and no feedback blocks (placebo–punishment, placebo–reward, placebo–no feedback, sulpiride–punishment, sulpiride–reward, sulpiride–no feedback). As we were mainly interested in the global differences between punishing and rewarding environments, we used every trial within these blocks regardless of the feedback that participants received. Probabilistic feedback was used within each block rather than 100% “punishment” or “reward,” as it was critical that participants thought their movement controlled the feedback they received. Although we perform analysis on the dominant trial type within each block, for example punishment trials during a punishment block, a lack of trials meant it was not feasible to investigate every trial type within each block.

Statistics. For the behavioral measures (MA, RT) in experiment 1, a repeated measures ANOVA (ANOVA-rm) compared session (placebo, sulpiride) and block (punishment, reward, no feedback) separately for mean and SD. Only MA was assessed within experiment 2. With MEP amplitude in experiments 1 and 2, an ANOVA-rm compared session (placebo, sulpiride) and block (punishment, reward, no feedback) separately for mean and SD. As the additional 12 participants tested in experiment 1 were not exposed to the no feedback block and had two TMS time points, separate statistics were performed. An ANOVA-rm compared session (placebo, sulpiride) and block (punishment, reward) separately for the mean and SD of MA. For MEP amplitude, an ANOVA-rm compared session (placebo, sulpiride), block (punishment, reward), and TMS time (150, 300) separately for mean and SD. In experiment 1, Pearson correlations were performed between the SD of MA and FDI/ADM for the punishment blocks across participants. Note that all 24 participants who had experienced the punishment block during the placebo and sulpiride session were used. For experiment 3, an ANOVA-rm compared session (placebo, sulpiride) and time points (T1, T2, T3) separately for the mean and SD of the MEP amplitude. Paired *t* tests were performed on significant interactions. The threshold for all statistical comparisons was $p < 0.05$. All data are mean \pm SEM.

Results

Psychological parameters

Participants felt significantly more fatigued during the sulpiride session relative to placebo (placebo, 3.0 ± 0.2 ; sulpiride, 3.8 ± 0.3 ; paired t test, $t_{(23)} = 2.6$, $p = 0.02$, two-tailed), with 12 of 24 participants correctly identifying the sulpiride and placebo sessions. However, there was no significant difference between sessions for the participant's rating of attention (placebo, 5.1 ± 0.2 ; sulpiride, 5.1 ± 0.3).

Experiment 1: behavior

The within-subject MA standard deviation (MA_{sd}) was significantly greater for punishment than reward and no feedback during placebo, but this effect was abolished by sulpiride (Fig. 2*b*).

For the mean of MA (MA_{mean}) there were no significant main effects or interaction for session (placebo, sulpiride) and block (punishment, reward, no feedback) (Table 1, Fig. 2*a*). In contrast, for MA_{sd} there was a significant main effect of session ($F_{(1,11)} = 5.8$, $p = 0.04$) and session \times block interaction ($F_{(2,22)} = 4.9$, $p = 0.02$), but no main effect of block. Paired t tests showed that MA_{sd} was significantly greater for placebo–punishment relative to sulpiride–punishment, placebo–reward, sulpiride–reward, placebo–no feedback, and sulpiride–no feedback ($t_{(11)} > 2.4$, $p < 0.03$, two-tailed; Fig. 2*b*). There were no significant differences between the other block types.

There was no significant difference between placebo–reward and placebo–no feedback for MA_{sd} ($t_{(11)} = 0.1$, $p = 0.9$; Fig. 2*b*). This might appear surprising as reward is often associated with a decrease in variability (Takikawa et al., 2002). However, in our experiment rewards were given randomly and so similar percentages of fast and slow movements (relative to the mean) were rewarded. As reward motivates the participant to repeat movements, this even distribution of reward across fast and slow movements would mean there was no necessity to decrease variability relative to baseline. In order for this to be true, the data should have a normal distribution with a similar amount of fast and slow MAs being associated with reward. To test this, the Shapiro–Wilk test for normality was performed for each participant on their raw MA data. Across all subjects, the test did not reach significance for both placebo–reward and placebo–no feedback ($0.44 < p > 0.06$), suggesting that MA in both block types maintained a normal distribution. Next, MA_{mean} within placebo–reward was used to separate fast and slow MAs for each participant. Thus, a fast MA was defined as any MA that was greater than the MA_{mean} , and a slow MA was any

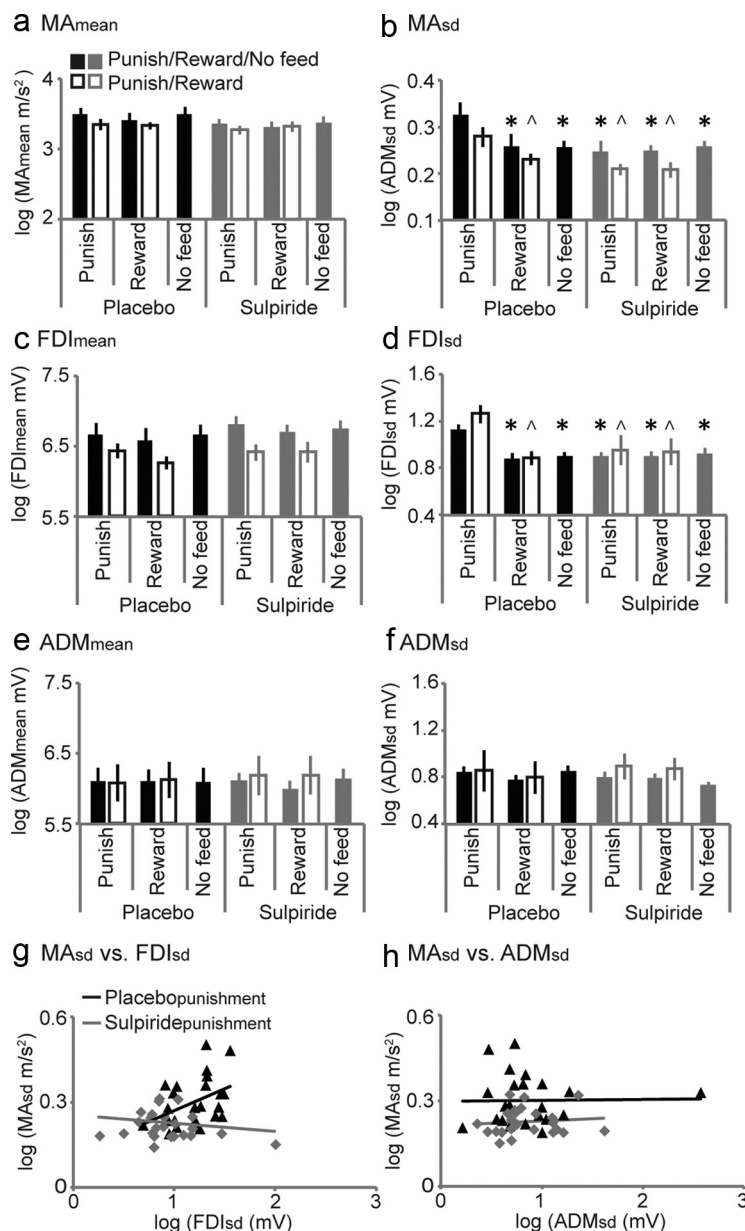


Figure 2. Experiment 1: greater behavioral and CSE variability during punishment is abolished by sulpiride. *a–f*, MA_{mean} (*a*), MA_{sd} (*b*) (log m/s \pm SEM), FDI_{mean} (*c*), FDI_{sd} (*d*), ADM_{mean} (*e*), and ADM_{sd} (*f*) (log mV \pm SEM). Data are shown for the punishment (Punish), reward (Reward), and no feedback (No feed) blocks during the placebo (solid black) and sulpiride (solid gray) sessions. Asterisk (*) indicates significant difference between placebo–punishment and placebo–reward, placebo–no feedback, sulpiride–punishment, sulpiride–reward, and sulpiride–no feedback ($p < 0.03$). There were no significant differences between reward and no feedback. The empty bars represent the data from the additional 12 participants who were only exposed to the punishment and reward blocks during placebo (empty black) and sulpiride (empty gray) sessions. Caret (^) indicates significant difference between placebo–punishment and placebo–reward, sulpiride–punishment, and sulpiride–reward ($p < 0.05$). *g–h*, Pearson correlation between MA_{sd} and FDI_{sd} (*g*) or MA_{sd} and ADM_{sd} (*h*) for placebo–punishment (black) and sulpiride–punishment (gray). Note that all 24 participants who had experienced the “punishment” block during the placebo and sulpiride session were used. Note also that all values have been multiplied by 1000 before log transformation.

MA that was less than the MA_{mean} . As reward feedback was provided on 80% of the trials, we then calculated the amount of fast and slow MAs that were rewarded for each participant within placebo–reward. In support of our conclusion, a paired t test revealed no significant difference between the amount of fast (21 ± 2 trials) and slow (19 ± 2) rewarded MAs across participants ($t_{(11)} = 0.5$, $p = 0.6$; two-tailed).

Similar results are observed if only the trials which pertain to that particular block type were used; for example punishment

Table 1. Experiment 1: Greater behavioral and CSE variability during punishment is abolished by sulpiride

	MA _{mean} (log m/s ²)	MA _{sd} (log m/s ²)*	RT _{mean} (log msec)	RT _{sd} (log msec)	FDI _{mean} (log mV)	FDI _{sd} (log mV)*	ADM _{mean} (log mV)	ADM _{sd} (log mV)
Placebo–punishment	3.50 ± 0.01	0.32 ± 0.03	5.14 ± 0.14	0.29 ± 0.04	6.66 ± 0.17	1.12 ± 0.05	6.09 ± 0.20	0.84 ± 0.05
Placebo–reward	3.41 ± 0.11	0.26 ± 0.03	5.06 ± 0.17	0.27 ± 0.04	6.59 ± 0.18	0.87 ± 0.05	6.10 ± 0.17	0.77 ± 0.05
Placebo–no feedback	3.50 ± 0.11	0.25 ± 0.02	5.10 ± 0.12	0.28 ± 0.04	6.67 ± 0.15	0.89 ± 0.04	6.08 ± 0.21	0.85 ± 0.05
Sulpiride–punishment	3.35 ± 0.08	0.24 ± 0.02	5.09 ± 0.15	0.29 ± 0.03	6.81 ± 0.11	0.90 ± 0.04	6.12 ± 0.11	0.80 ± 0.05
Sulpiride–reward	3.31 ± 0.08	0.25 ± 0.01	5.03 ± 0.15	0.25 ± 0.03	6.71 ± 0.10	0.89 ± 0.05	6.00 ± 0.12	0.79 ± 0.04
Sulpiride–no feedback	3.38 ± 0.09	0.26 ± 0.01	4.99 ± 0.12	0.27 ± 0.06	6.75 ± 0.12	0.92 ± 0.05	6.13 ± 0.15	0.73 ± 0.04

All data are mean ± SEM. Note that MA, FDI, and ADM values have been multiplied by 1000 prior to log transformation. Asterisk (*) indicates significant difference between placebo–punishment and all other conditions ($p < 0.05$).

trials during the punishment block (block: $F_{(1,11)} = 3.9$, $p = 0.03$; session: $F_{(1,11)} = 6.8$, $p = 0.02$; block × session: $F_{(1,11)} = 6.7$, $p = 0.005$; paired t tests; placebo–punishment vs sulpiride–punishment, placebo–reward, sulpiride–reward, placebo–no feedback, or sulpiride–no feedback; $t_{(11)} > 2.9$, $p < 0.01$, two-tailed). Although of interest, a lack of trials meant it was not feasible to investigate the other trial types within each block, for example the punishment trials during the reward block.

It is possible that an increase in SD could simply represent outliers. Before performing a log transformation, we ran separate Shapiro–Wilk tests of normal distribution on each participant's MAs within the punishment block for the placebo and sulpiride sessions. For all comparisons, the Shapiro–Wilk test was not significant ($0.35 < p > 0.05$). This suggests that SD is a valid measure of behavioral variability rather than a parameter distorted by outliers.

To reiterate, as the task instructions clearly stated that feedback was dependent on MA, we did not expect to observe any manipulation of RT across blocks or sessions. For the mean (RT_{mean}) and SD (RT_{sd}) of RT there were indeed no significant main effects or interaction for session and block (Table 1).

Experiment 1: CSE excitability

For the task-involved FDI, the SD of the MEP amplitude (FDI_{sd}) was significantly higher for punishment than reward and no feedback during placebo, but again this effect was abolished by sulpiride (Fig. 2*d*). During the placebo session, there was a positive correlation across participants between MA_{sd} and FDI_{sd} that was not present within the sulpiride session or in the task-noninvolved ADM (Fig. 2*g,h*).

There were no significant main effects or interactions for the FDI mean MEP amplitude (FDI_{mean} ; Table 1; Fig. 2*c*). For FDI_{sd} there was a significant main effect of block ($F_{(2,22)} = 4.7$, $p = 0.02$) and session × block interaction ($F_{(2,22)} = 6.7$, $p = 0.005$); however, the main effect of session was not significant. Paired t tests showed that FDI_{sd} was significantly greater for placebo–punishment relative to sulpiride–punishment, placebo–reward, sulpiride–reward, placebo–no feedback, and sulpiride–no feedback ($t_{(11)} > 2.7$, $p < 0.02$, 2-tailed; Fig. 2*d*). There were no significant differences between the other block types.

There were no significant main effects or interactions for the noninvolved ADM mean MEP amplitude (ADM_{mean} ; Table 1, Fig. 2*e*) or its variability (ADM_{sd} ; Table 1, Fig. 2*f*).

The muscles (FDI, ADM) were directly compared with an ANOVA-rm [session (placebo, sulpiride), block (reward, punishment, no feedback), muscle (FDI, ADM)]. FDI_{mean} was significantly larger than ADM_{mean} ($F_{(1,11)} = 40$, $p = 0.0005$), but no other effects were observed (Fig. 2*c,e*). In addition, we found a significant session × block × muscle interaction for FDI_{sd} versus ADM_{sd} ($F_{(2,22)} = 5.3$, $p = 0.01$). A paired t test revealed that FDI_{sd} was significantly larger than ADM_{sd} in the punishment block

during the placebo session ($t_{(11)} = 4.2$, $p = 0.002$; two-tailed; Fig. 2*d,f*).

Experiment 1: TMS timing

The additional 12 participants replicated the differences in variance between punishment and reward and showed that a TMS pulse at 150 or 300 ms after monetary feedback reflects similar changes in neurophysiological variability.

An ANOVA-rm compared session (placebo, sulpiride), block (punishment, reward), and, when appropriate, TMS time (150, 300). No effects on MA_{mean} , FDI_{mean} , ADM_{mean} , or ADM_{sd} were observed (Fig. 2). However, for MA variability (MA_{sd}) we found a main effect of session and interaction between session and block, but no effect of block (session: $F_{(1,11)} = 13$, $p = 0.005$; block × session: $F_{(1,11)} = 5$, $p = 0.04$). Paired t tests showed that MA_{sd} was significantly greater for placebo–punishment relative to sulpiride–punishment, placebo–reward, and sulpiride–reward ($t_{(11)} > 2.1$, $p < 0.05$, two-tailed; Fig. 2*b*). For FDI_{sd} there was a significant main effect of block ($F_{(1,11)} = 14$, $p = 0.003$) and block × session interaction ($F_{(1,11)} = 7$, $p = 0.03$). Importantly, all other main effects and interactions were not significant. As a result, the data were collapsed across TMS time. Paired t tests revealed that the FDI_{sd} for placebo–punishment was significantly greater than sulpiride–punishment, placebo–reward, and sulpiride–reward ($t_{(11)} > 2.2$, $p < 0.05$, two-tailed; Fig. 2*d*).

Experiment 1: correlation between behavioral and neurophysiological variability

During placebo–punishment there was a clear increase in variance for both the behavioral (MA) and neurophysiological (FDI) parameters. To assess whether these were associated, we performed a Pearson correlation between the MA_{sd} and FDI_{sd} for the punishment block across participants. Note that all 24 participants who had experienced the punishment block during the placebo and sulpiride session were used. During the placebo session, there was a significant positive correlation between these parameters ($r = 0.4$, $n = 24$, $p = 0.045$; two-tailed; Fig. 2*g*) that was not observed during the sulpiride session (Fig. 2*g*) or for the ADM muscle during either the placebo or sulpiride sessions (Fig. 2*h*).

Experiment 2: execution of predetermined MA

Punishment, reward, and sulpiride had no effect on behavioral or neurophysiological variability during a task where participants were required to execute a predetermined MA. For MA_{mean} , MA_{sd} , FDI_{mean} , FDI_{sd} , ADM_{mean} , and ADM_{sd} there were no significant main effects for session, block or interactions between session and block (Fig. 3*A*). A Pearson correlation was performed between MA_{sd} and FDI_{sd} for placebo–punishment across participants. Unlike experiment 1, there was no significant correlation between these parameters.

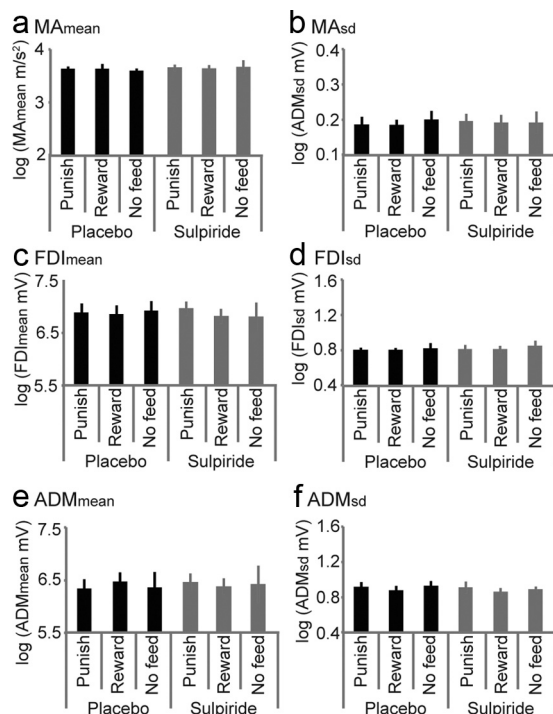


Figure 3. Experiment 3: execution of a predetermined MA is not affected by sulpiride. **a–f**, MA_{mean} (**a**), MA_{sd} (**b**) ($\log m/s \pm SEM$), FDI_{mean} (**c**), FDI_{sd} (**d**), ADM_{mean} (**e**), and ADM_{sd} (**f**) ($\log mV \pm SEM$). Data are shown for the punishment (Punish), reward (Reward), and no feedback (No feed) blocks during the placebo (black) and sulpiride (gray) sessions. Note that all values have been multiplied by 1000 before log transformation.

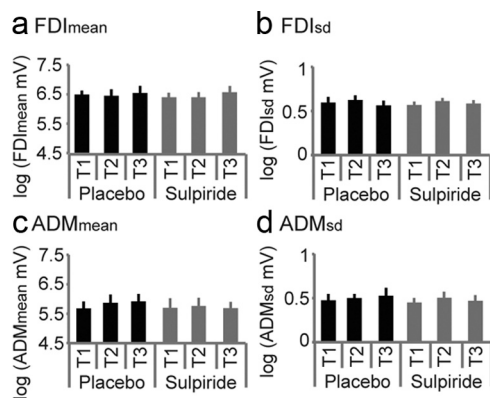


Figure 4. Experiment 4: resting CSE is not affected by sulpiride. **a–d**, FDI_{mean} (**a**), FDI_{sd} (**b**), ADM_{mean} (**c**), and ADM_{sd} (**d**) ($\log mV \pm SEM$). Data are shown for MEPs measured at rest either before (T1), in between (T2), or after (T3) experiments 2 and 3 during the placebo (black) and sulpiride (gray) sessions. Note that all values have been multiplied by 1000 before log transformation.

Experiment 3: influence of sulpiride on resting CSE excitability

There was no change in the mean or SD of the resting MEP amplitude when measured either before (T1), between (T2), or after (T3) experiments 2 and 3 in either the placebo or sulpiride sessions. For FDI_{mean} , FDI_{sd} , ADM_{mean} , and ADM_{sd} there were no significant main effects for session, time points, or interactions between session and time points (Fig. 4).

Discussion

This study tested whether low-level movement parameters are affected by DA-dependent processes of selection in ways similar

to those that operate during high-level action selection. During a task in which participants were required to select a MA, we found that blocks biased toward punishment were associated with increased MA variability relative to blocks of reward and no feedback. We show that this increase in behavioral variability was positively correlated with muscle-specific variability in CSE, suggesting a neurophysiological analog. Finally, we demonstrate that the administration of a D_2 antagonist caused the variability associated with punishment to diminish and the correlation between behavioral and neurophysiological variability to disappear. Similar changes in variability were not observed when participants were required to execute a predetermined MA, nor did DA influence CSE variability at rest.

Action selection is thought of as the high-level or cognitive process that selects between competing movements (Shmuelof and Krakauer, 2011). It has repeatedly been found that the decision of which action to select is influenced by DA and based on optimizing reward and minimizing cost/punishment (Schultz, 2006; Niv et al., 2007). For example, Pessiglione et al., (2006) showed that participants have a greater propensity to choose the most rewarding action after the administration of DA agonist. Frank et al., (2004) revealed that Parkinson's patients, who suffer from a deficit in DA, were better at learning to avoid negative outcomes than learning from positive outcomes, but DA medication reversed this bias. Learning which decision to make in any given situation can involve a process of trial and error (Fee and Goldberg, 2011). During the initial stages of learning, it is important that behavior is variable so that a sufficiently large task space is explored for the desired balance between cost and reward. DA is thought to be crucial for either inducing or shaping neural and behavioral variability during action selection (Costa, 2011; Fee and Goldberg, 2011). In mice, DA is important for inducing the novel activity patterns in cortico-Basal Ganglia that drives such motor exploration (Costa et al., 2006; Costa, 2011). Similarly, in pigeons the administration of a D_2 agonist during operant reinforcement learning increases behavioral variability (Pesek-Cotton et al., 2011), whereas a D_1 agonist has little effect (Ward et al., 2006). In contrast, a DA antagonist increased behavioral variability in songbirds (Leblois et al., 2010). Other work in songbirds suggests that rather than inducing variability, DA shapes neural and behavioral variability by providing a reinforcement signal that indicates good or bad song performance (Fee and Goldberg, 2011). Interestingly, DA also seems to control behavioral variability observed once an optimal action is found (Leblois and Perkel, 2012); however this is not the focus of the current paper. Crucially, in humans it was not previously known how behavioral variability is affected by rewarding or punishing outcomes and the relationship this has with DA.

There is now growing interest as to how DA-dependent processes not only influence the selection of an action but also its low-level kinematic parameters such as speed and force. Mazzoni et al. (2007) showed that Parkinson's patients implicitly select to move slower even though they preserve the ability to execute faster movement speeds with the same accuracy. This was not in the context of any explicit punishment or reward but suggests that patients select a different balance between the costs of moving fast and the rewards of completing the task and demonstrates that the selection of kinematic parameters can be manipulated by levels of DA (Mazzoni et al., 2007). Takikawa et al., (2002) showed that rewarding feedback is associated with decreased variability in saccadic velocity, latency, and amplitude, and Pessiglione et al., (2007) found that higher amounts of monetary reward lead to greater force being applied. Although this latter

work does not reveal a role for DA, it suggests that reward processes can have a direct impact on low-level movement parameters.

As mentioned previously, during trial-and-error learning, behavioral variability is important for the optimal action to be found (Fee and Goldberg, 2011), yet no study had directly investigated in humans whether there are DA-dependent changes in low-level movement parameter variability associated with punishment and reward. We found that during blocks of punishment the variability of MA increased, but this effect was abolished by a D₂ antagonist. This supports work in pigeons showing that D₂ receptors are important for behavioral variability (Pesk-Cotton et al., 2011). We propose that the punishment-induced increase in variability during the placebo session of experiment 1 reflects the participant's exploration of MA for a less punishing outcome. As the D₂ antagonist blocked this effect, we believe that this fits well with the role of D₂ receptors in the avoidance of negative or punishing behavioral outcomes (Frank et al., 2004; Kravitz et al., 2012). In addition, this result sides with the proposal that phasic DA shapes behavioral variability by providing a reinforcement signal that indicates performance outcome (Costa, 2011; Fee and Goldberg, 2011).

It is surprising that reward did not have the opposite effect to punishment and reduce movement variability in rewarded blocks, as in experiments on primates (Takikawa et al., 2002). However, in those experiments monkeys received a reward for eye movements in a particular direction, so that they knew on presentation of cue whether a trial was likely to receive reward. This may have motivated them to move faster and less variably on those trials. In the present experiments subjects did not know in advance whether a reward would accompany a movement. Additionally, with the present experimental design in which reward was given randomly, similar percentages of fast and slow movements were rewarded. This is similar for punishment, however; participants are motivated to avoid repeating the punished movement and therefore punishment of equally low and high MA increases variability. This is not the same for reward, where participants are motivated to repeat rewarded movements. Since a high proportion of movements were rewarded, there was little necessity to decrease variability relative to baseline. In retrospect, this is a limitation of the present study, and future work will attempt to address this issue.

As shown by Mazzoni et al. (2007), there appears to be a separation between selection (of a particular speed) and the execution of a movement (at high accuracy), with mildly affected Parkinson's disease patients being impaired in the former but not the latter. To examine whether the DA-dependent changes in variability were specific to selection rather than execution, we repeated the experiment but this time told participants to execute a predetermined MA. We found that punishment was now similar to conditions of reward and no feedback, and that administration of a D₂ antagonist did not affect variability across feedback type. Thus, although DA processes influence the selection of movement parameters, at least for the simple type of movements investigated here, they do not appear to manipulate their execution (Shmuelof and Krakauer, 2011). This fits well with models of motor control that propose independent neural loops for selection and execution of movement (Shadmehr and Krakauer, 2008; Izawa and Shadmehr, 2011) but do not exclude the possibility that DA processes may still be involved in the quality of action execution (Costa et al., 2004).

Recent work has suggested that the M1 might play an active role in the process of action selection (Cisek and Kalaska, 2010).

Here, we sought to test whether the variability of CSE may reflect changes in behavioral variability. In support of this, variability of CSE was greatest during punishment, i.e., when MA variability was highest. Critically, this was positively correlated with MA variability across participants. The increase in CSE variability and positive correlation were both abolished by a D₂ antagonist. No such relationship was observed when a predetermined MA was executed. This suggests that the relationship between behavioral and neurophysiological variability was dependent on the increased behavioral variability caused by punishment. At present, we are unsure as to the neural origin of such DA-dependent behavioral and neurophysiological variability. It is possible that the variability was originating from the basal ganglia, as D₂ antagonists are expressed abundantly there (Frank and O'Reilly, 2006). However, prior work has revealed that repetitive TMS over M1 leads to enhanced variability during performance (Teo et al., 2011), with the neurophysiological effects of TMS being impaired by a D₂ antagonist (Monte-Silva et al., 2011). In addition, it has been shown that DA-dependent changes in neural variability also occur in M1 (Costa et al., 2004; Costa et al., 2006). Yet, as there are reciprocal connections between the Basal Ganglia and M1 (Watabe-Uchida et al., 2012) it is difficult to disambiguate where DA is acting upon in the present design. Nevertheless, our results clearly show that CSE variability following punishing outcomes closely relates to behavioral variability.

In conclusion, DA-dependent processes of selection, which govern behavioral variability, appear to influence low-level movement parameters in ways similar to those that operate during high-level action selection. We propose that the enhanced behavioral variability associated with punishment reflects the participant's exploration of kinematic parameters for a less punishing or, conversely, a more rewarding outcome. This increased behavioral variability has a neurophysiological analog and is controlled via DA.

References

- Bestmann S, Harrison LM, Blankenburg F, Mars RB, Haggard P, Friston KJ, Rothwell JC (2008) Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Curr Biol* 18:775–780. [CrossRef Medline](#)
- Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* 33:269–298. [CrossRef Medline](#)
- Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M (2009) Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *J Neurosci* 29:1538–1543. [CrossRef Medline](#)
- Costa RM (2011) A selectionist account of de novo action learning. *Curr Opin Neurobiol* 21:579–586. [CrossRef Medline](#)
- Costa RM, Cohen D, Nicolelis MA (2004) Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr Biol* 14:1124–1134. [CrossRef Medline](#)
- Costa RM, Lin SC, Sotnikova TD, Cyr M, Gainetdinov RR, Caron MG, Nicolelis MA (2006) Rapid alterations in corticostriatal ensemble coordination during acute dopamine-dependent motor dysfunction. *Neuron* 52:359–369. [CrossRef Medline](#)
- Deleu D, Northway MG, Hanssens Y (2002) Clinical pharmacokinetic and pharmacodynamic properties of drugs used in the treatment of Parkinson's disease. *Clin Pharmacokinet* 41:261–309. [CrossRef Medline](#)
- Fee MS, Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 198:152–170. [CrossRef Medline](#)
- Frank MJ, O'Reilly RC (2006) A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behav Neurosci* 120:497–517. [CrossRef Medline](#)
- Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive

- reinforcement learning in parkinsonism. *Science* 306:1940–1943. [CrossRef Medline](#)
- Galea JM, Jayaram G, Ajagbe L, Celnik P (2009) Modulation of cerebellar excitability by polarity-specific noninvasive direct current stimulation. *J Neurosci* 29:9115–9122. [CrossRef Medline](#)
- Izawa J, Shadmehr R (2011) Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput Biol* 7:3. [CrossRef Medline](#)
- Kravitz AV, Tye LD, Kreitzer AC (2012) Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat Neurosci* 15:816–818. [CrossRef Medline](#)
- Leblois A, Perkel DJ (2012) Striatal dopamine modulates song spectral but not temporal features through D1 receptors. *Eur J Neurosci* 35:1771–1781. [CrossRef Medline](#)
- Leblois A, Wendel BJ, Perkel DJ (2010) Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioural variability through D1 receptors. *J Neurosci* 30:5730–5743. [CrossRef Medline](#)
- Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *J Neurosci* 27:7105–7116. [CrossRef Medline](#)
- Monte-Silva K, Ruge D, Teo JT, Paulus W, Rothwell JC, Nitsche MA (2011) D2 receptor block abolishes theta burst stimulation-induced neuroplasticity in the human motor cortex. *Neuropsychopharmacology* 36:2097–2102. [CrossRef Medline](#)
- Nitsche MA, Lampe C, Antal A, Liebetanz D, Lang N, Tergau F, Paulus W (2006) Dopaminergic modulation of long-lasting direct current-induced cortical excitability changes in the human motor cortex. *Eur J Neurosci* 23:1651–1657. [CrossRef Medline](#)
- Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* 191:507–520. [CrossRef Medline](#)
- Pesek-Cotton EF, Johnson JE, Newland MC (2011) Reinforcing behavioral variability: an analysis of dopamine-receptor subtypes and intermittent reinforcement. *Pharmacol Biochem Behav* 97:551–559. [CrossRef Medline](#)
- Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442:1042–1045. [CrossRef Medline](#)
- Pessiglione M, Schmidt L, Draganski B, Kalisch R, Lau H, Dolan RJ, Frith CD (2007) How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science* 316:904–906. [CrossRef Medline](#)
- Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? *Nat Rev Neurosci* 7:967–975. [CrossRef Medline](#)
- Schultz W (2006) Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol* 57:87–115. [CrossRef Medline](#)
- Schultz W (2007) Multiple dopamine functions at different time courses. *Annu Rev Neurosci* 30:259–288. [CrossRef Medline](#)
- Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. *Exp Brain Res* 185:359–381. [CrossRef Medline](#)
- Shmuelof L, Krakauer JW (2011) Are we ready for a natural history of motor learning? *Neuron* 72:469–476. [CrossRef Medline](#)
- Takikawa Y, Kawagoe R, Itoh H, Nakahara H, Hikosaka O (2002) Modulation of saccadic eye movements by predicted reward outcome. *Exp Brain Res* 142:284–291. [CrossRef Medline](#)
- Teo JT, Swayne OB, Cheeran B, Greenwood RJ, Rothwell JC (2011) Human θ burst stimulation enhances subsequent motor learning and increases performance variability. *Cereb Cortex* 21:1627–1638. [CrossRef Medline](#)
- Ward RD, Bailey EM, Odum AL (2006) Effects of D-amphetamine and ethanol on variable and repetitive key-peck sequences in pigeons. *J Exp Anal Behav* 86:285–305. [CrossRef Medline](#)
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74:858–873. [CrossRef Medline](#)